

**This Page Is Inserted by IFW Operations
and is not a part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

Best Available Copy

This Page Blank (uspto)

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-232694

(43)Date of publication of application : 02.09.1998

(51)Int.Cl. G10L 3/00
G10L 3/00

(21)Application number : 09-034575

(71)Applicant : MATSUSHITA ELECTRIC IND CO LTD

(22)Date of filing : 19.02.1997

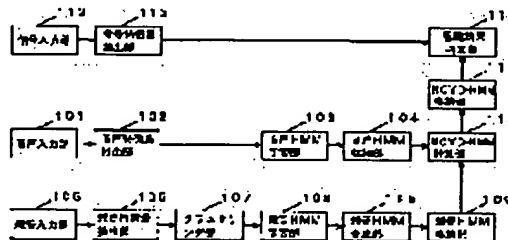
(72)Inventor : TANIGUCHI KENICHI
KONO NOBUYUKI
TOKUDA TADAMICHI
IKURA HIROO

(54) DEVICE AND METHOD FOR VOICE RECOGNITION

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a voice recognition device capable of precisely recognizing a voice that a noise is changed on the way of utterance in voice recognition under the noise by using a hidden Markov model.

SOLUTION: This device is provided with a clustering part 107 separating a non-stationary noise to plural stationary parts by power, a noise HMM(hidden Markov model) learning part 108 obtaining stationary noise HMMs for plural stationary noises and a noise HMM synthetic part 115 synthesizing a composite noise HMM from these stationary noise HMMs. Then, by NOVO (voice mixed with noise) converting a standard voice HMM by this composite noise HMM, a NOVO-HMM seasoned with this non-stationary noise is generated even when an environmental noise in a place where the voice recognition device is used is the non-stationary noise, and the voice is recognized even when the noise is changed on the way of utterance.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the
examiner's decision of rejection or application
converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of
rejection][Date of requesting appeal against examiner's decision
of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japanese Patent Office

This Page Blank (uspto)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-232694

(43) 公開日 平成10年(1998) 9月2日

(51) Int. Cl.⁵

G 1 0 L 3/00

識別記号

5 3 5

5 2 1

F I

G 1 0 L 3/00

5 3 5

5 2 1 L

審査請求 未請求 請求項の数 6 O L (全 9 頁)

(21) 出願番号

特願平9-34575

(22) 出願日

平成9年(1997) 2月19日

(71) 出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72) 発明者 谷口 賢一

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(72) 発明者 香野 信幸

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(72) 発明者 徳山 泰道

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

(74) 代理人 井理士 滝本 智之 (外1名)

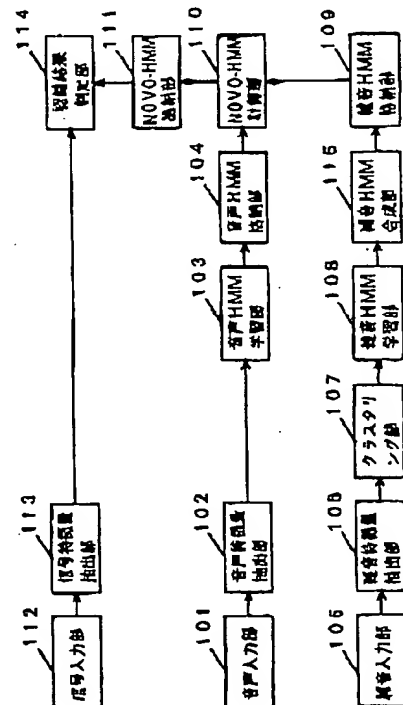
最終頁に続く

(54) 【発明の名称】 音声認識装置および音声認識方法

(57) 【要約】

【課題】 隠れマルコフモデルを用いた雑音下での音声認識において、発声途中に雑音に変化する音声を高い精度で認識することができる音声認識装置を提供することを目的とする。

【解決手段】 非定常雑音をパワーにより複数の定常部に分離するクラスタリング部107と、複数の定常雑音に対する定常雑音HMMを求める雑音HMM学習部108と、これら定常雑音HMMから1つの複合雑音HMMを合成する雑音HMM合成部115とを設け、この複合雑音HMMにより標準音声HMMをNOVO変換することにより、音声認識装置が使用される場所の環境雑音が非定常な雑音であっても、この非定常な雑音を加味したNOVO-HMMが生成され、発声途中に雑音に変化する場合でも音声を認識することができる。



【特許請求の範囲】

【請求項1】入力された音声および雑音信号をA/D変換する音声入力部と、所定間隔毎に入力信号を分割して所定間隔毎の周波数特徴量に分析する特徴量抽出部と、認識すべき単語の標準パターンが格納される音声HMM格納部と、入力する雑音をそのパワーにより複数の定常雑音成分に分離する雑音分離部と、前記雑音分離部によって得られた複数の定常雑音成分のそれぞれについて定常雑音HMMを求める定常雑音学習部と、複数の定常雑音HMMから1つの複合雑音HMMを合成する雑音合成部と、音声HMMと前記雑音合成部によって合成された複合雑音HMMをNOVO変換することにより雑音が重畳した音声のHMMを合成するNOVO-HMM計算部と、認識の対象となる音声信号の特徴量と前記NOVO-HMM計算部によって得られたHMMとを基に尤度を計算して最も尤もらしい単語を認識結果として決定する認識結果判定部とを備えたことを特徴とする音声認識装置。

【請求項2】入力する雑音を雑音のパワースペクトルにより定常雑音に近似できるように分離する雑音分離部を行することを特徴とする請求項1記載の音声認識装置。

【請求項3】入力された音声信号をA/D変換する音声入力部と、所定間隔毎に入力信号を分割して所定間隔毎の周波数特徴量に分析する特徴量抽出部と、認識すべき単語の標準パターンが格納される音声HMM格納部と、入力する雑音をそのパワーにより複数の定常雑音成分に分離する雑音分離部と、前記雑音分離部によって得られた複数の定常雑音成分のそれぞれについて定常雑音HMMを求める定常雑音学習部と、複数の定常雑音HMMから1つの複合雑音HMMを合成する雑音合成部と、前記雑音合成部によって得られた複合雑音HMMを格納する雑音HMM格納部と、標準パターンの標準音声HMMと前記雑音合成部によって合成された複合雑音HMMをNOVO変換することにより雑音が重畳した音声のHMMを合成するNOVO-HMM計算部とを備え、

初めに音声認識装置が使用される環境の雑音を収録し、前記雑音分離部における処理と、前記定常雑音学習部における処理と、前記雑音合成部における雑音合成処理を予め行って得られた複合雑音HMMを前記雑音HMM格納部に格納し、前記音声HMM格納部に格納された標準パターンの標準音声HMMと前記雑音HMM格納部に格納された複合雑音HMMとを基に前記NOVO-HMM計算部にてNOVO変換して雑音が重畳したHMMを合成し、認識の対象となる音声信号が入力されると、その音声信号の特徴量と前記NOVO-HMM計算部により得られたHMMとを基に尤度を計算して最も尤もらしい単語を認識結果として決定することを特徴とする音声認識装置。

【請求項4】HMMを用いて雑音が重畳された音声認識する音声認識方式であって、初めに音声認識装置が使

用される環境の雑音を収録し、この雑音をそのパワーにより複数の定常雑音成分に分離し、得られた複数の定常雑音成分のそれぞれについて定常雑音HMMを求め、これら複数の定常雑音HMMから1つの複合雑音HMMを合成し、標準パターンの標準音声HMMと前記複合雑音HMMとを基に雑音が重畳したNOVO-HMMを合成し、認識の対象となる音声信号が入力されると、その音声信号の特徴量と前記NOVO-HMMとを基に尤度を計算して最も尤もらしい単語を認識結果として決定することを特徴とする音声認識方法。

【請求項5】HMMを用いて雑音が重畳された音声認識する音声認識方式であって、認識に用いるNOVO-HMMを作成するために、標準音声HMMと雑音HMMを合成する時に、音声と雑音との平均残差パワーの比を係数として重み付けすることを特徴とする音声認識方法。

【請求項6】HMMを用いて雑音が重畳された音声認識する音声認識方式であって、認識に用いるNOVO-HMMを作成するために、予め任意に重み係数を与えて仮のNOVO-HMMを作成し評価する処理を重み係数を変えながら繰り返すことで、認識率の良い重みの係数を求めておき、実際に認識に用いるNOVO-HMMを作成する場合には、前記予め求めた係数を用いて重み付けすることを特徴とする音声認識方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明はHMM方式を用いた音声認識装置および音声認識方法に関するものである。

【0002】

【従来の技術】計算機による音声の自動認識に広く用いられている手法に隠れマルコフモデル (Hidden Markov Model: 以下、本発明ではHMMと称す) によるものがある。初めにHMMによる音声認識の方法について説明する。

【0003】HMMは、N個の状態 S_1, S_2, \dots, S_N を持ち、一定間隔毎に、ある確率 (遷移確率) で状態を次々に遷移するとともに、その際に、ある確率 (出力確率) でラベル (特徴データ) を一つずつ出力する。音声をラベル (特徴データ) の時系列と見た場合に、学習時に、各単語を数回発声してそれらをモデル化したHMMを作成しておく、そして、未知の入力音声を認識する時には、その入力音声のラベル系列に合致したラベル系列を出力する確率が最大であるHMMを探し、そのHMMに対応する単語を出力結果とする。この手法は最尤推定法と呼ばれている。

【0004】さらに詳細に説明すると、学習時に認識対象の人の音声サンプル群と、認識対象の各単語毎にHMMを準備しておく、そしてこれらのHMMがその認識対象の音声サンプル群から抽出される特徴データ系列を出力し易いようにそのHMMを定義している内部パラメー

タを調節する、この際forward-backwardアルゴリズムを用いてHMMの内部パラメータを調節し、各HMMにはその認識対象の単語にマッチした内部パラメータが設定される。

【0005】未知の音声が入力されると、最尤推定法により、各HMM毎にその未知の音声から抽出した特徴データ系列の出力し易さ(尤度)を算出し、最大の尤度を出力したHMMに対応する単語を認識結果とする。

【0006】このように予め単語毎にそのHMMを学習して、各単語に最も適した状態の遷移確率と各状態遷移におけるラベルの出力確率を求めておけば、未知の単語のラベル系列が入力した時に各HMMに対して確率(尤度)計算を行えば、どの単語に対するHMMがこのラベル系列を出力し易いかわかり、これにより認識ができる。

【0007】HMMを用いて雑音が重畳された音声进行認識する手法の一つにFranc Martinが文献「Recognition of Noisy Speech by Composition of HMM (信学技報SP92-96)」で提案したNOVO-HMMを用いる方法がある。これは雑音から作成したHMMすなわち「雑音HMM」と「標準パターンの音声HMM」の内部パラメータを、前記文献中でNOVO (Voice mixed with noise) 変換と呼ばれている手法で合成し、こうして作成された「雑音が重畳された音声HMM」すなわちNOVO-HMMを用いることにより、雑音が重畳された音声を高い精度で認識するというものである。

【0008】図9はNOVO変換の概念図であり、認識対象単語の学習サンプルデータを用いた学習によって標準音声HMMを生成し、雑音の学習サンプルデータを用いた学習によって雑音HMMを生成し、これら標準音声HMMと雑音HMMとをNOVO変換によって合成し、各認識対象単語毎にNOVO-HMMを得る。

【0009】図6は、従来のNOVO-HMMが表す対数スペクトルの概形図であり、図7は雑音重畳音声入力して作成したHMMが表す対数スペクトルの概形図である。両者で大体同じはずである概形が異なることが分かる。結果として認識率の低下を招いていた。

【0010】図10は従来のNOVO変換におけるHMMの内部パラメータの計算手順のフローチャートである。従来の方式のNOVO変換では、まず、標準音声HMMおよび雑音HMMの内部パラメータであるケプストラムをCOS変換によって対数スペクトルへ変換する(step1)。

【0011】次に、どちらも指数変換を行なって線形スペクトルに変換する(step2)。その後、2つの線形スペクトルを加算し、標準音声と雑音とを重ねたものの線形スペクトルを作成する(step3)。そして、作成した線形スペクトルを対数変換によって対数ス

ペクトルに戻す(step4)。さらに逆COS変換する(step5)ことにより、標準音声と雑音の重畳したもののケプストラムを得る。

【0012】2つの線形スペクトルの加算の部分の計算式は、Franc Martinの文献のHMMcompositionの項に記述されているように、以下の(数1)、(数2)のようになる。

【0013】

【数1】

$$\mu^{s1n} = \mu^{s1n} + k(SNR) \times \mu^{n1n}$$

【0014】

【数2】

$$\Sigma^{s1n} = \Sigma^{s1n} + k^2(SNR) \times \Sigma^{n1n}$$

【0015】ここで、k(SNR)は次の(数3)のように表されます。

【0016】

【数3】

$$k(SNR) = \sqrt{\frac{S_{pow}}{N_{pow}}} \cdot 10^{\frac{SNR}{20}}$$

【0017】なお、以上の数式中の μ は平均ベクトル、 Σ は分散の行列である。また R_{1n} 、 S_{1n} 、 N_{1n} は各々、雑音重畳音声、音声、雑音を意味する。SNRは雑音重畳時のSN比である。 S_{pow} と N_{pow} は、各々HMMの学習に用いた音声および雑音のパワーの平均値である。

【0018】そして、この(数3)でのk(SNR)は、雑音が重畳された音声のSN比により変わる、つまり雑音のパワーのみで種類には無関係のパラメータである。例えば、学習時に用いる音声と雑音のパワーを等しくしておき、SN比が0dB(SNR=0)となるように雑音を重畳した音声を認識する場合、k(SNR)の値は雑音の種類によらず1となる。

【0019】

【発明が解決しようとする課題】しかしながら、以上のNOVO-HMMによる認識手法でも雑音が重畳された音声を認識すると概ね良好な結果が得られていたが、それは雑音が発声時間中に大きく変化しないことが必要であり、発声途中で雑音の種類が大きく変化した場合には認識率が大きく低下してしまうという問題点があった。

【0020】例えば図9に示す従来のNOVO変換では、雑音を重畳した音声を認識する場合、k(SNR)の値は雑音の種類によらず1となる。また従来の認識方式では標準パターンの音声HMMと「雑音HMM」を、雑音の種類が何であっても同じように合成するため、雑音の影響が大きくなる場合、そのNOVO-HMMの表現では対応できていない。

【0021】本発明は、隠れマルコフモデルを用いた非定常な雑音下での音声認識において、発声途中で雑音が

変化する音声を高い精度で認識することができる音声認識装置および音声認識方法を提供することを目的とする。

【0022】

【課題を解決するための手段】本発明の音声認識装置においては、入力する雑音をそのパワーにより複数の定常雑音成分に分離する雑音分離部と、この雑音分離部によって得られた複数の定常雑音成分のそれぞれについて定常雑音HMMを求める定常雑音学習部と、複数の定常雑音HMMから1つの複合雑音HMMを合成する雑音合成部とを設け、この複数の定常雑音HMMから合成して得た複合雑音HMMと標準パターンの音声HMMとをNOVO変換するように構成したので、発声途中で雑音に変化する場合でも音声を高い精度で認識することができる。

【0023】また、標準パターンの音声HMMと雑音HMMとを合成する過程で、音声の平均残差パワーと雑音の平均残差パワー（雑音の予測値パワーと実際のパワーとの差の平均値）の比を係数として重み付けをし、雑音が重畳された音声を表現するNOVO-HMMを作成する（請求項5）。

【0024】あるいは、予め、任意に重み係数を与えて仮のNOVO-HMMを作成し評価することを、重み係数を変えながら繰り返すことで、認識率の良い重みの係数を求めておき、その予め求めた係数で重み付けをし、雑音が重畳された音声を表現するNOVO-HMMを作成する（請求項6）。

【0025】

【発明の実施の形態】本発明の請求項1に記載の発明は、入力された音がおよび雑音信号をA/D変換する音声入力部と、所定間隔毎に入力信号を分割して所定間隔毎の周波数特徴量に分析する特徴量抽出部と、認識すべき単語の原型パターンが格納される音声HMM格納部と、入力する雑音をそのパワーにより複数の定常雑音成分に分離する雑音分離部と、雑音分離部によって得られた複数の定常雑音成分のそれぞれについて定常雑音HMMを求める定常雑音学習部と、複数の定常雑音HMMから1つの複合雑音HMMを合成する雑音合成部と、音声HMMと雑音合成部によって合成された複合雑音HMMをNOVO変換をすることにより雑音が重畳した音声のHMMを合成するNOVO-HMM計算部と、認識の対象となる音声信号の特徴量とNOVO-HMM計算部によって得られたHMMとを基に尤度を計算して最も尤もらしい単語を認識結果として決定する認識結果判定部とを備えたものであり、これにより、音声認識装置が使用される場所の環境雑音が非定常な雑音であっても、この雑音に似たスペクトルの雑音でクラスタリングして複数の定常的な雑音の成分に分離し、この複数の定常雑音成分のそれぞれについて雑音HMMを作成し、これら複数の雑音HMMを合成することによって複合な雑音を加味

したNOVO-HMMが生成され、非定常な雑音を加味したNOVO変換をすることが可能となる。

【0026】本発明の請求項5に記載の発明は、HMMを用いて雑音が重畳された音声を認識する音声認識方式であって、認識に用いるNOVO-HMMを作成するために、標準音声HMMと雑音HMMを合成する時に、音声と雑音との平均残差パワーの比を係数として重み付けすることを特徴とする音声認識方法であり、これにより、雑音が重畳された音声を認識するNOVO-HMMの作成に際して、標準パターンの音声HMMと雑音HMMとの合成の重みを考慮し、雑音の種類によらず、最高の認識精度が出せるNOVO-HMMを作成することができる。

【0027】以下、本発明の一実施の形態による音声認識装置および音声認識方法について図面を参照しながら説明する。

【0028】（実施の形態1）図1は本発明の一実施の形態による音声認識装置の構成ブロック図である。101は標準パターンを作成するための音声信号をデジタル値に変換する音声入力部、102は標準パターン音声信号からフレーム毎に特徴量を算出する音声特徴量抽出部、103は複数の標準パターン音声特徴量から各単語毎に標準パターンとなるHMM（以下、標準音声HMMとする）を作成する音声HMM学習部、104は標準音声HMMを格納する音声HMM格納部である。

【0029】105は雑音標準パターンを作成するための雑音信号をデジタル値に変換する雑音入力部、106は雑音信号からフレーム毎に特徴量を算出する雑音特徴量抽出部である。107は雑音特徴量から似た分析データを集めて、クラスタリングする雑音分離部としてのクラスタリング部である。108はクラスタリングされた雑音特徴量から、それぞれの雑音毎に標準パターンとなる雑音HMMを作成する雑音HMM学習部である。115はそれぞれの雑音HMMを1つの雑音HMMに合成する雑音HMM合成部である。109は1つに合成された雑音HMMを格納する雑音HMM格納部である。

【0030】110は標準音声HMMと雑音HMMを合成して雑音が重畳した音声のHMMをNOVO法により合成するNOVO-HMM計算部、111は雑音が重畳した標準音声HMMを格納するNOVO-HMM格納部である。

【0031】112は音声認識の対象となる音声信号をデジタル値に変換する信号入力部、113は入力信号からフレーム毎に特徴量を算出する信号特徴量抽出部、114は入力単語の出力確率を計算すると共に認識結果の決定を行う認識結果判定部である。

【0032】図2は本発明の一実施の形態による音声認識装置の回路ブロック図である。201は音声を電気信号に変換するマイク（マイクロホン）、202は中央処理装置（CPU）、203は読み出し専用メモリ（ROM）

M)、204は書き込み可能メモリ(RAM)、205は出力装置である。図1の構成ブロック図における信号入力部112および音声入力部101は、マイク201とCPU202により構成される。また図1における各特徴量抽出部とデータ格納部と認識結果判定部114は、CPU202がROM203に格納されたプログラムを実行し、RAM204にアクセスすることにより実行される。

【0033】図1における音声入力部101および雑音入力部105および信号入力部112はマイク201とCPU202により構成されている。また音声特徴量抽出部102と音声学習HMM部103と雑音特徴量抽出部106とクラustering部107と雑音学習HMM部108と雑音HMM合成部115と信号特徴量抽出部113とNOVO-HMM計算部110と認識結果判定部114は、CPU202がROM203に格納されたプログラムを実行し、RAM204にアクセスすることにより実行される。このとき、音声HMM格納部104および雑音HMM格納部109およびNOVO-HMM格納部はRAM204により構成される。

【0034】図3は本実施の形態による音声認識方法のフローチャートである。まず、音声入力部101から標準パターンとなる音声を入力する。一標準パターンあたり数十から数百人分の音声波形を収集し、入力とする(step1)。その音声波形の音声区間に対し、LPC(Liner Predictive Coding)ケプストラム分析などの分析方法で周波数分析を行なう(step2)。これらの音声周波数分析データを基に、不特定話者用の標準パターンとなる標準音声HMMをforward-backwardアルゴリズムを用いて作成する(step3)。

【0035】次に、雑音入力部105から音声認識装置が使用される場所の非定常な環境雑音を入力する(step4)。その雑音信号に対し、LPCケプストラム分析などの分析方法で周波数分析を行なう(step5)。雑音信号をパワーの基準により似た雑音信号に分離する(step6)。あるいはパワースペクトルのスペクトル距離で分離することができる。例えば、この際、分離基準としてはLPCケプストラム距離尺度を用いることができる。これらの雑音周波数分析データを基に、分離された各クラスターの雑音を連結して、それぞれの雑音の標準パターンとなる雑音HMMをforward-backwardアルゴリズムを用いて作成する(step7)。

【0036】そして、それぞれの雑音HMMを後述する雑音HMM合成方法で1つの雑音HMMに合成し、合成された雑音HMMを雑音HMM格納部109に格納する。そして音声HMM格納部104上に格納されている各標準音声毎に雑音HMM格納部109で格納されている各雑音HMMとNOVO変換を行なう(step

8)。

【0037】未知の音声が入力されると、次のように尤度計算を行い、最大の尤度を出力したNOVO-HMMに対応する単語名を認識結果として出力装置205に出力する。具体的には、マイク201に音声が入力されると(step9)、信号特徴量抽出部113を経て得られた特徴量をRAM204に書き込む(step10)。次に、ROM203上に格納されている各標準音声毎に作成したNOVO-HMMについて、RAM204上の特徴量に対する尤度を計算する(step11)。そして、最大の尤度を出力したNOVO-HMMに対応する単語名を認識結果として出力装置15に出力する(step12)。

【0038】通常、標準音声HMMは音声認識の応用される目的により予め決まった単語の音声波形データを収録し(step1)、音声特徴量の抽出(step2)、標準音声HMMの計算(step3)まで実行しておいて、標準パターンとなるHMMをROM203に格納しておくことになる。

【0039】また、音声認識装置が使用される目的が単一で、音声認識装置が使用される環境雑音が変わらなければ、使用環境下の雑音データを予め収録し(step4)、雑音特徴量の抽出(step5)、クラustering(step6)、雑音HMMの計算(step7)、雑音HMMの合成(step13)、標準音声HMMと雑音HMMの合成(step8)まで実行しておいて、合成済のHMMをROM203に格納しておくこともできる。

【0040】図4は本発明の一実施の形態による音声認識方法に用いられるNOVO変換の概念図であって、NOVO-HMMの作成過程を示すものであり、図14に示す従来方法と比べると学習サンプルデータを用いた学習によって生成する雑音HMMの形状が異なる。

【0041】NOVO変換を施す雑音HMMの状態数と認識に際して考慮する雑音の種類数を共に2とした場合、従来の方法では1種類の雑音から2状態の雑音HMMを学習によって直接生成した後、その雑音HMMと単語HMMに対してNOVO変換を施していたが、本発明では、雑音HMM学習部によって変動する雑音から学習した複数個の雑音HMMを用いる。

【0042】例えば、雑音HMM学習部において2種類の雑音が学習されたとする。2種類の雑音から学習によってそれぞれ1状態の雑音HMMを生成し、状態遷移確率を人為的に与えることによって(自己遷移確率を0.7程度、他状態遷移確率を0.3程度)それら雑音HMMの状態を結合し、2状態の雑音HMMを生成する。そして、この2状態の雑音HMMと標準音声HMMとに従来のNOVO変換を施してNOVO-HMMを作成する。

【0043】このように本発明によれば、変動する雑音

に対し雑音重畳音声を高い確率で認識することができる。例えば自動車内においては、非定常雑音が、数100ms程度の定常区間の連続からなる非定常雑音とパルス状の重畳雑音であることから、定常区間の連続である非定常雑音に対し効果のある雑音対策方法で非定常雑音環境下での認識性能の向上を図ることができる。

【0044】（実施の形態2）図14に示す従来のNOVO変換は、雑音の学習サンプルデータを用いた学習によって雑音HMMを作成した後、これら標準音声HMMと雑音HMMとをNOVO変換によって合成し、各認識対象単語毎にNOVO-HMMを得ているが、それだけでは図9に示すように雑音を重畳した音声を認識する場合、 k (SNR) の値は雑音の種類によらず1となる。

【0045】しかし現実には、同じSN比でも雑音の種類によって平均残差パワーが異なり、平均残差パワーによってその雑音の影響度が違う。例えば比較的周回性の高い成分からなる雑音の平均残差パワーは低く、その場合は音声認識に対する雑音の影響は小さい。それに対し周回性の低い成分からなる雑音の平均残差パワーは大きく、このような平均残差パワーの大きい雑音ほど、雑音の影響が大きくなる。図7がこれに相当し、図6の既形との間に差がある。

【0046】本実施の形態は、雑音HMM合成部115にて各雑音の雑音HMMから1つの雑音HMMを合成する際、それぞれの雑音HMMにおいて標準音声の平均残差パワーとその雑音の平均残差パワーの比によって重み付けを行ない、各雑音HMMに個別の重み付け係数を掛けて合成する音声認識方法である。

【0047】図5は本発明の一実施の形態による音声認識方式に用いられるNOVO変換におけるHMMの内部パラメータの計算手順のフローチャートである。NOV

$$\mu^{r1n} = \mu^{s1n} + m \times k \text{ (SNR)} \times \mu^{n1n}$$

【0053】

【数5】

$$\Sigma^{r1n} = \Sigma^{s1n} + m^2 \times k^2 \text{ (SNR)} \times \Sigma^{n1n}$$

【0054】ここで、以上の数式における係数 m は次の（数6）または（数7）のように表される。

【0055】

【数6】

$$m = \sqrt{\frac{N_{\text{residual-pow}}}{S_{\text{residual-pow}}}}$$

【0056】

【数7】

$$m = (\text{評価の繰り返しにより求めた係数})$$

【0057】数式中の $N_{\text{residual-pow}}$ と $S_{\text{residual-pow}}$ は各々HMMの学習に用いた雑音および音声の残差パワーの平均値である。その他の意味は、前出の（数1）、（数2）、（数3）の説明と同様である。

O変換では、まず、標準音声HMMおよび雑音HMMの内部パラメータであるケプストラムをCOS変換によって対数スペクトルへ変換する（step1）。

【0048】次に、どちらも指数変換を行なって線形スペクトルに変換する（step2）。そして、標準音声の平均残差パワーと雑音の平均残差パワーを各々求める。これらは、各HMMの内部パラメータに残差パワーの項が含まれているため容易に求められる。次に標準音声の平均残差パワーと雑音の平均残差パワーの比によって重み付けをし、2つの線形スペクトルを加算する（step3）。

【0049】特許請求の範囲の請求項5に記載の発明は、標準音声の平均残差パワーと雑音の平均残差パワーの比を係数として、この係数を掛けることにより重み付けをし、2つの線形スペクトルを加算する。また請求項6に記載の発明は、予め、任意に重みを与えて仮のNOVO-HMMを作成し評価する処理を重みを変えながら繰り返すことで、認識率の良い重みの値を求めておき、その予め求めた値を係数として重み付けをし、2つの線形スペクトルを加算する。step3ではこれら2つの内の何れかの方法により線形スペクトルを加算する。

【0050】これらにより標準音声と雑音の重畳したものの線形スペクトルを作成する。そして、作成した線形スペクトルを、対数変換し（step4）、逆COS変換する（step5）ことにより標準音声と雑音の重畳したもののケプストラムを得る。

【0051】従来の技術の項にて示した2つの線形スペクトルの加算の部分の計算式（数1）および（数2）は、下記の（数4）および（数5）ように変形される。

【0052】

【数4】

【数5】

【0058】図8は、本発明の一実施の形態による音声認識方式により作成したNOVO-HMMを表す対数スペクトルの概形図であり、図7の、雑音重畳音声を入力して求めたHMMを表す対数スペクトルの概形図に近いことが分かる。このように本実施の形態は、雑音の種類によらず、雑音が重畳された音声を図8が示すようにうまく表現でき、結果として雑音重畳音声を高い確率で認識することができる。

【0059】

【発明の効果】以上のように本発明は、入力する雑音をそのパワーにより複数の定常雑音成分に分離する雑音分離部と、この雑音分離部によって得られた複数の定常雑音成分のそれぞれについて定常雑音HMMを求める定常雑音学習部と、複数の定常雑音HMMから1つの非定常

雑音HMMを合成する雑音合成部とを設け、この複数の定常雑音HMMから合成して得た複合雑音HMMと標準パターンの音声HMMとをNOVO変換するように構成したので、音声認識装置が使用される場所の環境雑音が非定常な雑音であっても、この非定常な雑音を加味したNOVO-HMMが生成され、発声途中に雑音に変化しても高い精度で認識することができる。

【0060】また標準音声HMMと雑音HMMを合成する時に、音声と雑音との平均残差パワーの比を係数として重み付けすることにより、重畳される雑音の種類によらず高い精度で認識する

【図面の簡単な説明】

【図1】本発明の一実施の形態による音声認識装置の構成ブロック図

【図2】本発明の一実施の形態による音声認識装置の回路ブロック図

【図3】本発明の一実施の形態による音声認識方法のフローチャート

【図4】本発明の一実施の形態による音声認識方法に用いられるNOVO変換の概念図

【図5】本発明の一実施の形態による音声認識方式に用いられるNOVO変換の計算手順のフローチャート

【図6】従来のNOVO-HMMから求めた対数スペクトルの概形図

【図7】雑音重畳音声を入力として作成したHMMから求めた対数スペクトルの概形図

【図8】本発明の一実施の形態による音声認識方式によ

り作成したNOVO-HMMが表す対数スペクトルの概形図

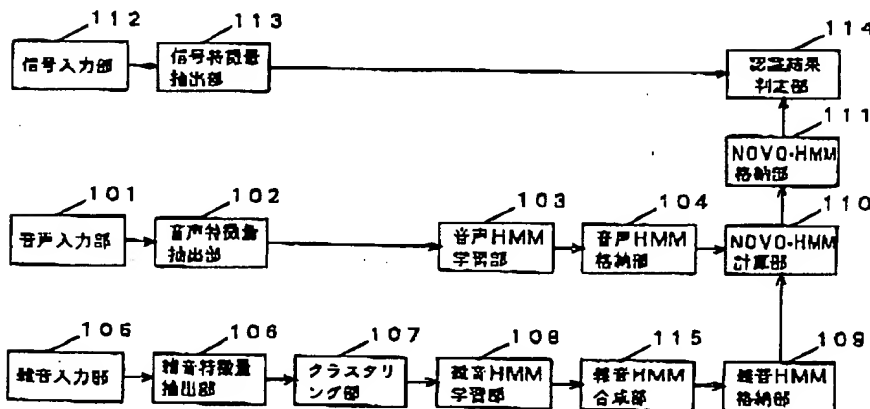
【図9】従来のNOVO変換の概念図

【図10】従来のNOVO変換におけるHMMの内部パラメータの計算手順のフローチャート

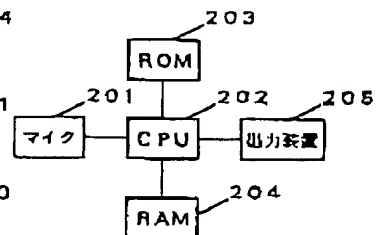
【符号の説明】

- 101 音声入力部
- 102 音声特徴量抽出部
- 103 音声HMM学習部
- 104 音声HMM格納部
- 105 雑音入力部
- 106 雑音特徴量抽出部
- 107 クラスタリング部
- 108 雑音HMM学習部
- 109 雑音HMM格納部
- 110 NOVO-HMM計算部
- 111 NOVO-HMM格納部
- 112 信号入力部
- 113 信号特徴量抽出部
- 114 認識結果判定部
- 201 マイク
- 202 CPU
- 203 ROM
- 204 RAM
- 205 出力装置

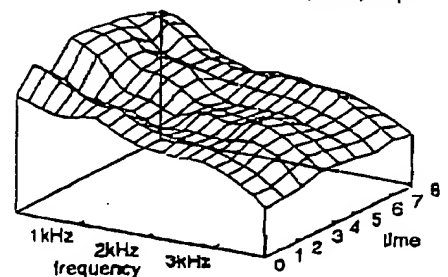
【図1】



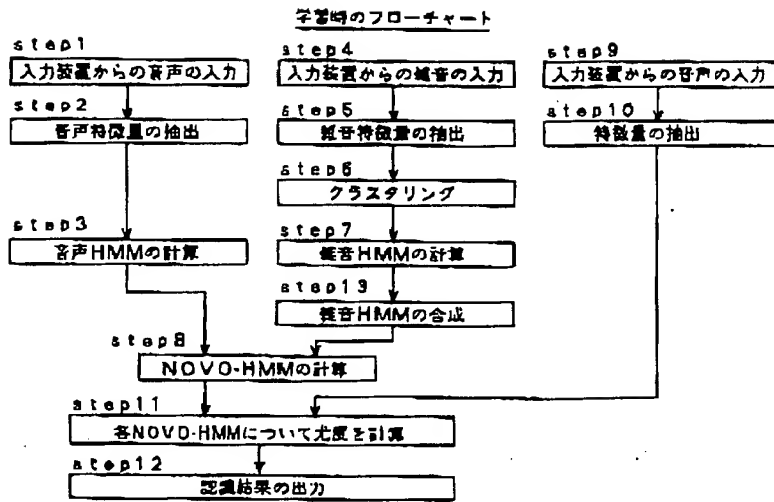
【図2】



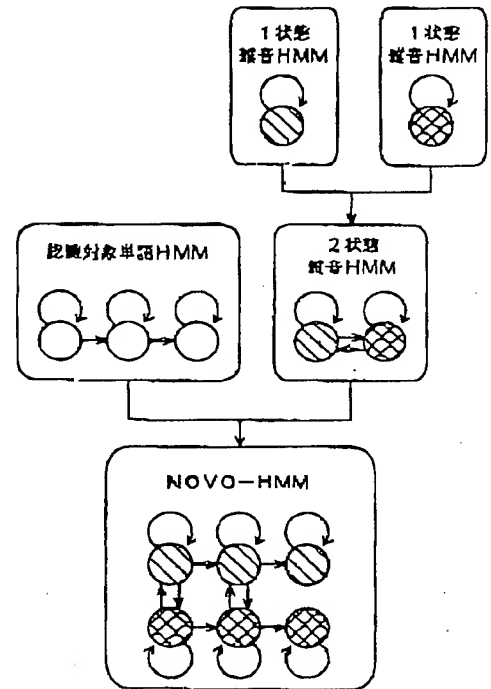
【図6】



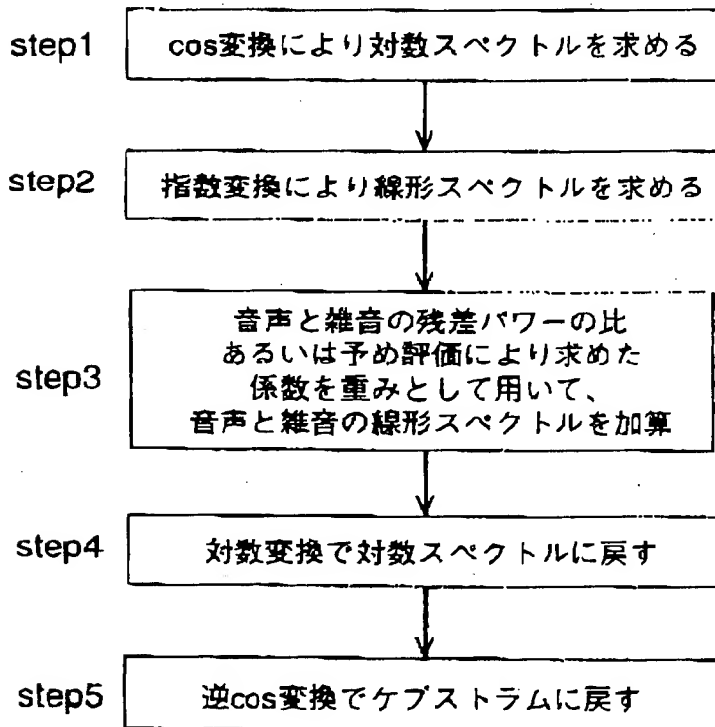
【図3】



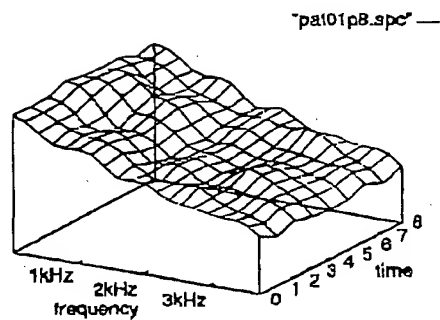
【図4】



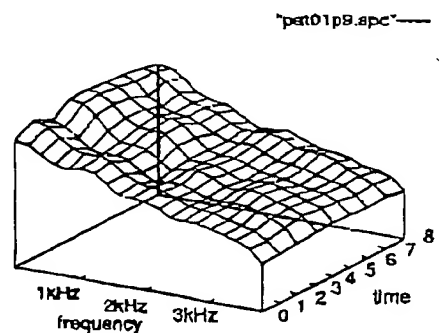
【図5】



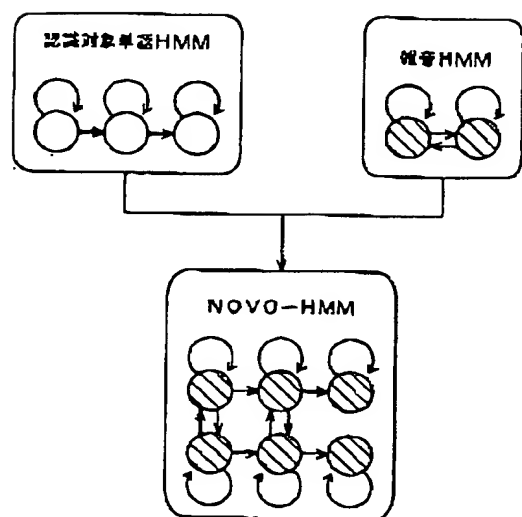
【図7】



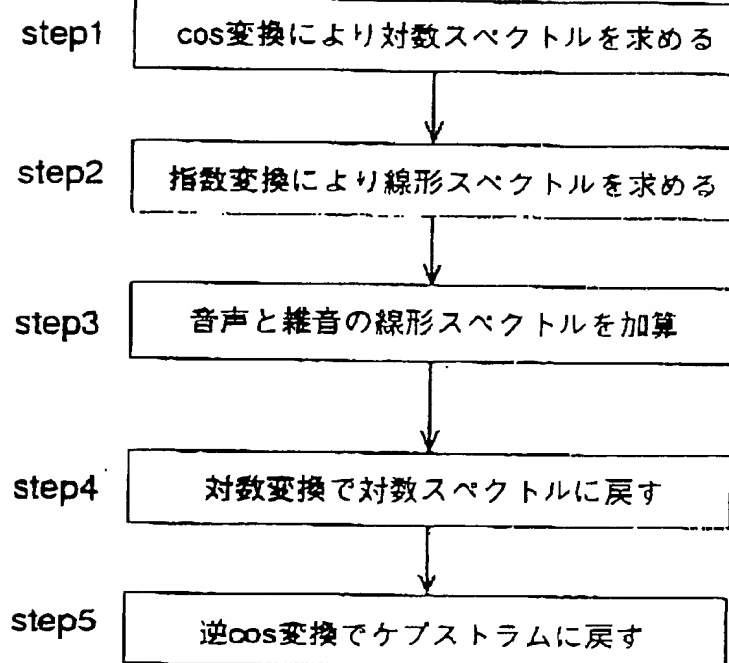
【図8】



【図9】



【図10】



フロントページの続き

(72)発明者 居食 啓雄
大阪府門真市大字門真1006番地 松下電器
産業株式会社内

This Page Blank (uspto)